

CUANTIFICACION VECTORIAL ADAPTATIVA DE LA VOZ

José Adrián Rodríguez Fonollosa, Rafael M. Carbonell y Enrique Masgrau*

ABSTRACT

Vector quantization (VQ) is a simultaneous quantization of a sequence of samples or vector. This process allows to make effective use of the interrelations among the different vector components and performance arbitrarily close to the ultimate rate-distortion can be achieved by VQ if the vector dimension is high enough. Nevertheless, the exponential growth in complexity forces the use of low -dimensionality VQ in practical systems. In this case some kind of adaptation is necessary to obtain adequate performance. Adaptive Multistage VQ (AMSVQ) [1] is a simple method that allows to redesign the quantizer as it is used, thus increasing its robustness and performance. In this paper, it is shown the close relation between the proposed adaptation algorithm and the methods usually employed to design the vector quantizers. Then some results are presented in which the AMSVQ reduce the complexity and improve the performance of previous VQ-based schemes of quantizing the LPC parameters.

INTRODUCCION

La cuantificación vectorial es una técnica que permite aprovechar al máximo la interrelación existente entre las muestras que forman el vector a cuantificar. Sin embargo, para obtener una codificación eficiente suele ser necesario introducir algún tipo de adaptación que aproveche las dependencias entre distintos vectores. En concreto, la señal de voz es un proceso localmente estacionario con una estadística que varía lentamente y un codificador, para ser eficiente, habrá de ser capaz de aprovechar esta estacionariedad. Es también un resultado conocido que, dada una velocidad de codificación, un cuantificador vectorial dependiente del locutor, esto es, diseñado específicamente para el locutor cuya voz se está codificando, trabaja mejor que un codificador independiente del locutor.

Por tanto, un cuantificador vectorial que se adapte a la señal a cuantificar al mismo tiempo que es usado daría lugar a un sistema eficiente y robusto frente a cambios de locutor, variaciones en la condiciones ambientales o en el equipo de adquisición de la señal de voz. Esta idea no es nueva y, por ejemplo, en [2] se describe un sistema que modifica las palabras código durante su uso. Sin embargo, este cambio crea la necesidad de transmitir los nuevos vectores código generados al receptor y esto supone un incremento significativo en la velocidad de codificación. Otros sistemas, aunque sí tienen un funcionamiento "backward", (sin transmisión adicional), adaptan únicamente la ganancia, como el cuantificador vectorial de ganancia adaptativa [3]. El esquema que aquí se presenta tiene la novedad de que se adapta tanto a la forma como a la ganancia de los vectores y no necesita para ello transmisión lateral adicional.

Este esquema está basado en la estructura multietapa, y por ello se le ha denominado Cuantificador Vectorial Multietapa Adaptativo, AMSVQ. El hecho de que las palabras código sean continuamente actualizadas da lugar a unos resultados más independientes del cuantificador diseñado inicialmente y esto supone, además, que las secuencias de entrenamiento no han de ser necesariamente tan largas y representativas como era necesario en los esquemas no adaptativos, ya que el cuantificador vectorial tiende a rediseñarse a si mismo.

Sin embargo, la ventaja más importante de la cuantificación multietapa adaptativa es que permite superar de forma sencilla uno de los más graves inconvenientes de la cuantificación vectorial que es su falta de robustez frente a cambios en la señal a codificar o en las condiciones de adquisición de la misma.

* Departamento de Teoría de la Señal y Comunicaciones. Universidad Politécnica de Cataluña.

This work was supported by the PRONTIC grant nº 105/88

Estas propiedades la hacen interesante en aplicaciones como la cuantificación de los coeficientes LPC y los resultados obtenidos muestran efectivamente se produce un incremento significativo en la calidad de la codificación respecto a esquemas no adaptativos.

ALGORITMO DE ADAPTACION

La cuantificación vectorial multietapa se ha tomado siempre como un sistema subóptimo con un coste computacional y de memoria reducido. A continuación veremos que la estructura multietapa puede aprovecharse, además, para obtener un sistema adaptativo que haga frente a cambios en la estadística de la señal a codificar y mejore las prestaciones del cuantificador multietapa convencional (no adaptativo).

En el cuantificador vectorial multietapa [4], el vector de entrada es aproximado sucesivamente por medio de varios cuantificadores vectoriales en cascada, donde el vector de entrada a cada etapa es el error de cuantificación de la etapa anterior.

El algoritmo de adaptación desarrollado aprovecha esta estructura para tratar de mejorar una etapa en función de la información suministrada por las siguientes etapas. En la figura 1 puede observarse el funcionamiento de un cuantificador vectorial de dos etapas. El vector de entrada $x(n)$ es cuantificado para obtener el vector de salida $z(n)$, resultado de sumar la salida de primer cuantificador $c_i(n)$ y la salida del segundo cuantificador $e_q(n)$.

$$z(n) = c_i(n) + e_q(n) \quad (1)$$

$$c_i(n) = q_1(x(n)) \quad (2)$$

$$e(n) = x(n) - c_i(n) \quad (3)$$

$$e_q(n) = q_2(e(n)) \quad (4)$$

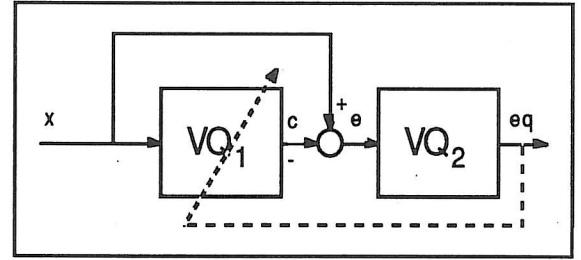


Figura 1. Cuantificador Vectorial Multietapa Adaptativo

Al cuantificar $x(n)$ con el primer cuantificador vectorial se produce un error $e(n)$, que es a continuación cuantificado por la segunda etapa. Nuestro objetivo al obtener el algoritmo de adaptación será el de minimizar el error de cuantificación de la primera etapa. Si la segunda etapa está correctamente diseñada tendremos como resultado una reducción en el ruido de cuantificación total.

En el diseño de los cuantificadores vectoriales se calcula la palabra código c_i o centroide como aquel vector que minimiza la distorsión sobre todos los vectores de entrenamiento que el codificador le ha asignado. En nuestro caso estamos buscando un sistema que se adapte a la nueva estadística y por ello es lógico dar un peso mayor a vectores más recientes. En concreto el objetivo será obtener la palabra código $c_i(n+1)$ que minimice el error cuadrático ponderado exponencialmente sobre los vectores que se le han asignado, esto es

$$E = \sum_{j=0}^{\infty} \beta^j \|x_i(n-j) - c_i(n+1)\|^2 \quad (5)$$

donde $x_i(n)$ hace referencia a los vectores cuantificados por la palabra código i -ésima. Derivando (5) respecto a $c_i(n+1)$ e igualando a 0 se obtiene

$$c_i(n+1) = \frac{\sum_{j=0}^{\infty} \beta^j x_i(n-j)}{\sum_{j=0}^{\infty} \beta^j} \quad (6)$$

$$c_i(n+1) = (1-\beta) \sum_{j=0}^{\infty} \beta^j x_i(n-j) \quad (7)$$

$$c_i(n+1) = (1-\beta) \sum_{j=1}^{\infty} \beta^j x_i(n-j) + (1-\beta) x(n) \quad (8)$$

$$c_i(n+1) = \beta c_i(n) + (1-\beta) x(n) = c_i(n) + (1-\beta) e(n) \quad (9)$$

Sin embargo este algoritmo, tal como aparece en (9) no es útil ya que el receptor no dispone del valor de $e(n)$ y estábamos buscando una adaptación "backward", que no incrementase la velocidad de codificación. El receptor, sin embargo, si que dispone de $e_q(n)$, que no es más que una estimación de $e(n)$, y se puede, por tanto, utilizar este valor para obtener la siguiente ecuación de actualización de la palabra código que será la que se utilizará tanto por el receptor como por el emisor.

$$c_i(n+1) = c_i(n) + \mu e_q(n) \quad (10)$$

donde se hecho además el cambio $\mu=(1-\beta)$.

Se puede observa claramente la relación entre (6) y el cálculo del centroide, y podemos decir, por tanto, que el algoritmo de adaptación es equivalente a un algoritmo LBG donde se da mayor peso a los vectores más recientes.

Coste computacional. Mediante este algoritmo sólo se modifica una palabra código cada vez que se realiza una cuantificación y por tanto, su coste computacional es independiente del tamaño del cuantificador y es de tan solo una multiplicación-suma por muestra, (N multiplicaciones-sumas por vector).

Diseño. El diseño del cuantificador vectorial multietapa adaptativo es semejante al del no adaptativo aunque para obtener la secuencia de entrenamiento de la segunda etapa hemos de tener en cuenta el efecto de la adaptación sobre la primera [1].

RESULTADOS

Uno de los primeros usos del cuantificador vectorial y del cuantificador vectorial multietapa fue la cuantificación de los coeficientes LPC. Los primeros trabajos en este área se centraron en obtener una codificación eficiente de estos parámetros en el contexto de los vocoders, cuya calidad venia limitada por su misma estructura. Estos resultados previos no se pueden aplicar directamente a los más recientes codificadores de alta calidad como pueden ser el codificador CELP [6], donde la cuantificación de los coeficientes LPC ha de ser mucho más precisa. Es por ello que el tema ha resurgido con fuerza en los últimos años y se siguen buscando métodos eficientes y robustos para la cuantificación de los coeficientes LPC.

Generalmente no se utiliza una cuantificación directa de los coeficientes LPC sino que se realiza una transformación previa de los mismos. Se busca con ello tener parámetros más independientes entre si y con una sensibilidad más uniforme que la que tienen los coeficientes LPC originales. En el caso de utilizar cuantificación vectorial las opciones más corrientes emplean como distancia el cociente de verosimilitud (likelihood-ratio), o una distancia cuadrática sobre los cocientes de log-áreas. La deducción del algoritmo de adaptación se ha realizado utilizando una distancia cuadrática y por ello se ha escogido una distancia de este tipo sobre los cocientes de log-área para las pruebas realizadas. En concreto la SNR sobre los cocientes de log-área v_i se define como

$$\text{LAR-SNR} = \frac{\sum_{m=1}^M \sum_{i=1}^N v_i^2(m)}{\sum_{m=1}^M \sum_{i=1}^N (v_i(m) - v_{qi}(m))^2}$$

donde M es número de tramas total y N el orden del predictor.

En las pruebas realizadas que se presentan a continuación se estudia en primer lugar la cuantificación de los coeficientes LPC utilizando la LAR-SNR como medida espectral de distorsión. Posteriormente se aplica esta cuantificación a un codificador CELP.

Base de datos. Para diseñar los cuantificadores vectoriales se utilizan una base de datos formada por 80 frases de 14 locutores diferentes, mientras que para las pruebas se utilizan 12 frases de otros 6 locutores diferentes. En el cálculo de los coeficientes LPC se utiliza el algoritmo de Levison-Durbin sin preénfasis y con una ventana de Hamming de 200 muestras (25 ms). El intervalo de cálculo escogido es de 160 muestras (20 ms) y se emplean 10 coeficientes LPC.

Cuantificación con 27 bits. Aun usando cuantificación vectorial suelen ser necesarios entre 25 y 30 bits por trama para obtener una alta calidad en la cuantificación de los parámetros LPC. El objetivo de este primer esquema de 27 bits por trama es el de obtener una codificación prácticamente transparente de estos parámetros y comparar estos resultados con los obtenidos con otros métodos. El número de etapas es 4; la primera tiene un tamaño 8 (3 bits) y las 3 etapas siguientes constan de 256 palabras código cada una. El número total de bits por trama es por tanto de 27, (3+8+8+8), que a una velocidad de 50 tramas por segundo dan una velocidad de codificación de los coeficientes LPC de 1350 bps.

En [7] se describe un sistema de codificación de los coeficientes LPC que utiliza también un cuantificador multietapa adaptativo. La adaptación se realiza, sin embargo de manera muy diferente, ya que se basa en la matriz de covarianza de los vectores anteriores y requiere además del cálculo de esta matriz, el cálculo de sus autovectores. Sin embargo, los resultados obtenidos con este esquema son útiles como comparación con los aquí obtenidos ya que utiliza la misma medida de distancia espectral. En la tabla I puede verse un resumen de la prestaciones obtenidas con el esquema de Atal, el cuantificador multietapa no adaptativo y el adaptativo. Se observa como el cuantificador multietapa adaptativo aquí propuesto permite superar el problema de la robustez y obtener resultados equivalentes a cuantificadores adaptativos mucho más complejos.

Esquema	bits/trama	Coste Comput./trama	SNR(dB)
[Atal 1987]	28	13056	20,50
MSVQ	27	776	17,07
AMSVQ	27	776	19,75

Tabla I. Comparación entre el esquema propuesto en [Atal 1987], el cuantificador multietapa no adaptativo y el adaptativo.

Cuantificación con 10 bits. Este ejemplo permite comparar el cuantificador vectorial multietapa adaptativo con el cuantificador de búsqueda completa. Este último es el cuantificador sin memoria óptimo, pero tal como puede observarse en la figura 3, donde se ha representado la relación señal a ruido obtenida con el cuantificador de búsqueda completa y varios cuantificadores de 2 etapas adaptativos da peores prestaciones que el esquema propuesto debido a que el algoritmo de adaptación del AMSVQ permite aprovechar la correlación entre vectores a corto y largo término.

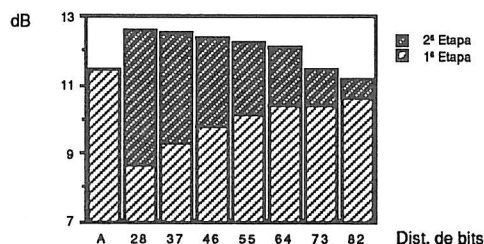


Figura 3. Comparación entre un cuantificador vectorial de búsqueda completa (A) y varios cuantificadores multietapa adaptativos de 10 bits.

En la tabla II se compara igualmente la complejidad y la relación señal a ruido entre los mejores esquemas adaptativos y el cuantificador vectorial de búsqueda completa.

Esquema	bits/trama	Coste Comput./trama	SNR(dB)
VQ	10	1024	11,52
AMSVQ	3+7	136	12,53
AMSVQ	2+8	260	12,62

Tabla II. Comparación entre el cuantificador de búsqueda completa y dos cuantificadores multietapa adaptativos de 10 bits.

Codificación de la señal de voz. Los cuantificadores diseñados han sido aplicados a la cuantificación de los parámetros LPC utilizados por un codificador CELP y un codificador multipulso. Los resultados muestran que, como era de esperar, el incremento en la LAR-SNR que permitía obtener el cuantificador multietapa adaptativo se traduce en incremento tanto de la SNR como de la calidad de la señal decodificada. El codificador CELP utilizado en las simulaciones tiene todos los otros parámetros sin cuantificar, (predictor largo y ganancia), y utiliza el cuantificador usual de dimensión 40 y tamaño 1024. Los resultados medios obtenidos al utilizar la cuantificación vectorial multietapa adaptativa son los que pueden verse en la siguiente tabla.

Esquema-bits	Sin Cuant.	AMSVQ-27	MSVQ-27	AMSVQ-19	MSVQ-19
SNR	10,57	10,00	9,65	9,38	9,15

CONCLUSIONES

En este artículo se ha mostrado que la estructura multietapa, además de suponer una reducción substancial del coste computacional y de memoria, permite la introducción de un algoritmo de adaptación. Este algoritmo de adaptación rediseña el cuantificador continuamente sin apenas coste computacional y ninguna transmisión adicional de bits. El resultado es un incremento notable en la robustez frente a cambios de locutor, idioma, condiciones ambientales o cualquier otro aspecto que modifique la estadística de la señal a cuantificar.

La cuantificación de los coeficientes LPC tiene gran importancia en la mayoría de los codificadores de voz a media y baja velocidad. De los estudios y resultados obtenidos se ha comprobado que el cuantificador vectorial multietapa adaptativo permite efectivamente incrementar notablemente la robustez y eficiencia en la cuantificación de estos parámetros. Esta mejora es importante en esquemas a velocidades inferiores a los 8 Kbps donde una codificación precisa y eficiente es fundamental para obtener una voz decodificada sin distorsión.

REFERENCIAS

- [1] "Adaptive Multistage Vector Quantization" José A. RODRIGUEZ FONOLLOSA, Enrique MASGRAU GOMEZ. IEEE Proc. MELECON, Lisboa, Portugal, April 1989.
- [2] "An 800 bps Adaptive Vector Quantization Vocoder using a perceptual distance measure". Douglas B. PAUL. IEEE Proc. ICASSP, No. 02.04, p. 73-76, April 1982.
- [3] "Gain Adaptive Vector Quantization with application to Speech Coding". Juin-Hwey CHEN, Allen GERSHO. IEEE Trans. on COM, No. 9, Vol. 35, p. 918-930, September 1987.
- [4] "Multiple stage Vector Quantization for Speech Coding". Biing H. JUANG, Augustine H. GRAY Jr.. IEEE Proc. ICASSP, p. 597-600, November 1982.
- [5] "An algorithm for Vector Quantizer design". Yoseph LINDE, Andrés BUZO, Robert M. GRAY. IEEE Trans. on COM, No. 1, Vol. 28, p. 84-95, January 1980.
- [6] "Code-excited linear prediction (CELP): High-quality Speech at very low bit rates". Proc. ICASSP, Vol. 3, p 941-944, March 1985.
- [7] "Stochastic Gaussian Model for Low-Bit Rate Coding of LPC Area Parameters". Bishnu S. ATAL. IEEE Proc. ICASSP, paper 51.1, p 614-617 1987.